Do word graphs stimulate design?

N. M. Segers, B. de Vries and *H. H. Achten*, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

In the early phase of the architectural design process the architect makes his ideas explicit in the design draft by means of a variety of representations such as sketches, marks, images, and annotations. So far, annotations have not seen much research attention, although they do reveal part of the design content. In this article we study the effects of offering feedback to annotations, namely by presenting word graphs that contain the architect's annotations and semantic associations based on these words. Our aim is to investigate whether the feedback in word graphs can stimulate design. The research method entails a working prototype, called the Idea Space System (ISS), in an experimental setting in which 18 architects participated. ISS searches for semantic relations between single words (annotations) and for intermediary words. The findings of the experiment show that architects appreciate the use of word graphs. They experience pleasure, an improvement of the workflow, and especially get more associations while using the ISS with feedback in word graphs. Nevertheless, we cannot find a significant increased creativity of the resulting designs, or a significant reduction of fixation.

Keywords: architectural design, computer aided design, conceptual design, creativity, design method

Design is characterised by complexity and ill-definedness or ill-structuredness (Simon, 1973) of design problems. Designers have to avoid premature commitment to a solution of a design problem, also known as fixation (Purcell and Gero, 1996). Given the lack of conclusive solutions, designing requires both logic and creativity (Alexander, 1964; Archer, 1969; Christiaans, 1992) to iteratively develop task and solution during the design process (Alexander, 1964; Dorst, 1997). The design is gradually developed through a series of sketches, images, annotations, marks, and so forth. These design representations together make up the design draft. They are tenuous, uncertain, sketchy, and personal. For support of the architect in the early design phase, CAAD systems should be able to respond to the design draft (Segers, 2004).

Corresponding author:
B. de Vries
b.d.vries@bwk.tue.nl

www.elsevier.com/locate/destud
0142-694X $ - see front matter *Design Studies* **26** (2005) 625–647
doi:10.1016/j.destud.2005.05.002

Sketches have been under much consideration in research (Goldschmidt, 1991; Achten, 1997; Do, 2002), while words have not been sufficiently paid attention to. This lack of attention seems undeserved as words have multiple ways of contributing to the design process. Explicitly, words can be uttered, written down as text, or written down in the design draft. We define *annotation* as any word that the architect writes down in the design draft. Annotations are used as comments, questions, descriptions, explanations, expressions, additions, and diagrams (Segers, 2004).

During the early phase of the design process words are constantly instilled with subtly different meanings and changing interpretations. Due to this quality of words, annotations have similar purposes as sketches (Lawson and Loke, 1997; Suwa et al., 1998; Segers, 2004). They serve as external memory; they provide visual cues for association of verbalised issues, and they serve as a physical setting in which thoughts are constructed on the fly in a situated way.

Association, as one of the creative cognitive processes (Finke et al., 1992) or creativity-relevant skills or processes (Amabile, 1983, 1996), is generally identified as an important mechanism to acquire new ideas. Offering associations even seems to increase the number of other kinds of associations (Silberman et al., 2001). We acquired similar results in a small design case study, in which we analysed the use of annotations (Segers, 2004). It appeared that words and their related associations helped in both diverging and converging ideas in the design process.

Based on this, we hypothesise that offering semantic associations on the basis of annotations will stimulate design, in the number of ideas, the manner in which ideas come together again, the logic in ideas and process, and the consistency of ideas. In order to test the influence of associations in the design process we chose to create a prototype design system in an experimental setting. A prototype allows us to provide similar conditions for each subject, to instantaneously prompt associations on an annotation, to display associations as feedback in one glance, and to combine all this in real-time with the design draft as the subject is designing. We can test our assumptions and determine whether our hypothesis holds, and if so, how relevant it is for architects in practice.

The outline of the article is as follows: firstly, we describe our research methodology of testing through design tasks with a prototype system in an experimental setting; secondly, the design of the experiment will be

described; thirdly, the results are discussed. The article ends with a review of the research results and the research methodology.

1 Research method

The prototype design system allows us to measure the effect of associations on architects as they design. There are many means of measuring effects on design: verbal reports, self-reports or protocol analysis, interview methods, observation, simulation and multi-dimensional techniques (Cordingly, 1989; Ericsson and Simon, 1993). In design research it is common practice to test assumptions or theories by means of case studies and experiments (Schön, 1983; Rowe, 1987; Valkenburg, 2000; Reymen, 2001). Creativity is commonly measured in terms of number of ideas, quality of ideas, and sometimes a combination of those (van der Lugt, 2001). Also, creativity is measured by a group of appropriate observers who are able to judge whether a product is creative (Amabile, 1983, 1996). Gero and Purcell (1996) measure fixation in creative processes by comparing a general list of features with a list of features of each design, produced by the experimental and control groups.

1.1 Prototype system: idea space system

The Idea Space System (ISS) (Segers and de Vries, 2003; Segers, 2004) is a prototype design system which captures the design draft, generates semantic associations, and optionally displays these in the form of so-called word graphs to the architect (Figure 1). The system has a horizontal workspace like a drawing table, and a vertical screen to display word graphs. The pen-based desk has a workbook User Interface metaphor. The strokes made with the pen are traced immediately as the architect makes them on the work desk (Aliakseyeu et al., 2001). The word graphs are projected on the vertical screen to prevent distraction of the architect during the design process and to avoid clutter on the work area.

Semantic associations are generated by the natural language ontology WordNet (Miller et al., 1990). WordNet is an online lexical reference system based on psycholinguistic theories of human lexical memory. WordNet includes nouns, verbs, adjectives and adverbs. The semantic associations of WordNet implemented in ISS are synonym, antonym, hypernym, hyponym, holonym, meronym, troponym, pertainym, similarity, and entailment.

Word graphs are visualised by ISS on the basis of the DOT library by Koutsofios and North (1993). DOT creates hierarchical layouts of

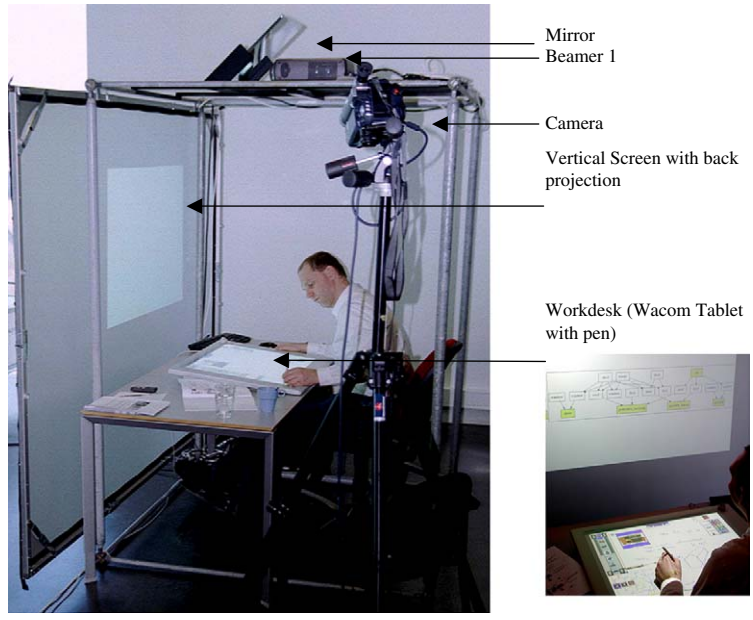

Figure 1 Set-up of ISS prototype during the experiment

directed graphs in a visually clear and simple manner: single words (as nodes) and arrows (representing edges to connect the words). Word graphs show in one glance the annotation and semantic associations of the annotation (see Figure 2).

The prototype saves the design draft as screenshot. Screenshots are made every 5 s, when a word graph is inserted, and when the architect turns to a next page. Figure 3 shows an example of a screenshot.

The screenshots document the design process as it develops through the design draft. The accepted word graphs actually become part of the

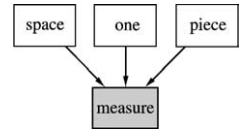

Figure 2 Example of a DOT graph, presenting three nouns the architect has written down in his design draft and one newly generated noun by the ISS, which relates the words above semantically by means of a hypernym association

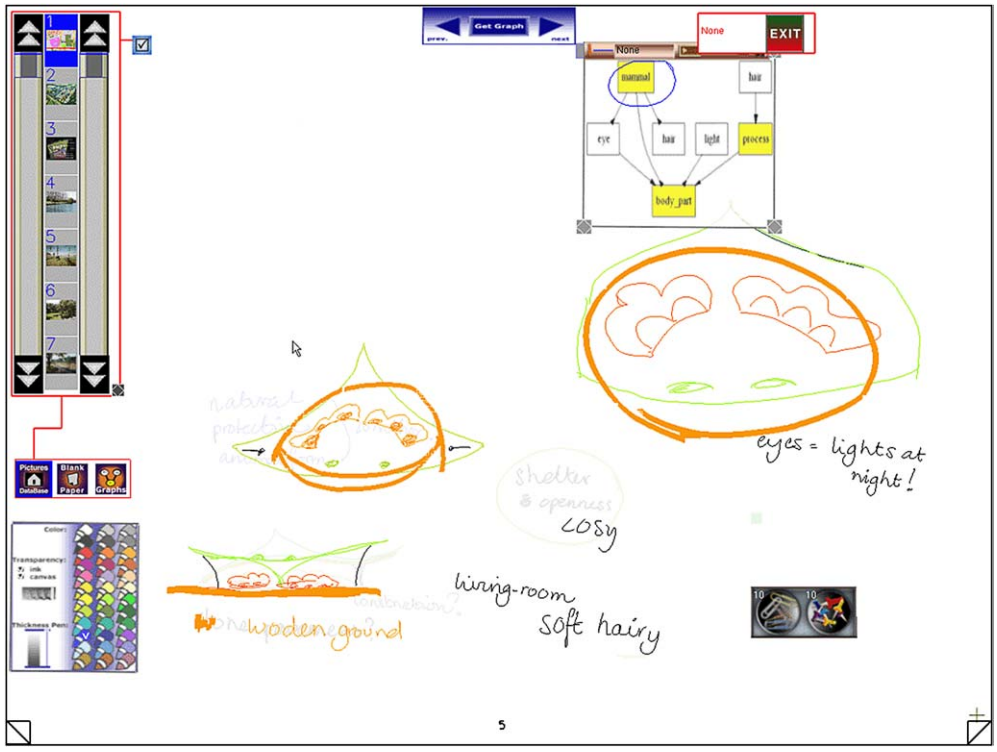

Figure 3 Example of design draft including word graph on work desk, architect 8, time: 39:28 (mm:ss)

design draft. Figure 4 shows in more detail the word graph of Figure 3. The design process can be described as follows:

The architect was searching for a comfortable and "cosy" space as a "living-room." She was thinking in materials: wood ("wooden ground") and some "soft, hairy" material. She was sketching benches in the plan view of the bus stop. The plan view sketch shows her eye-like features. So she writes down "eyes." "Eye" and "hair" are both part of a "mammal," which is a newly generated word of ISS. The architect marks this as being important for the design process. She is making

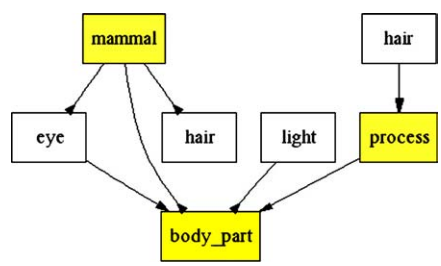

Figure 4 Word graph from Figure 3, enlarged

an analogy between the bus stop and a mammal. Somewhat later in the design process she will write down "fish": resulting from "mammal" (via animal) and the shape of the sketch.

It can be seen in the design draft that sketches, word graphs, and words contribute in making associations on the design. In this example we may infer that the word graphs do stimulate the design.

1.2 Design of the experiment

The experiment is a two-condition (with and without feedback in word graphs) repeated-measures design with two independent variables (task and order of the tasks). The subjects of the experiment are practising architects. Each subject had to make two design tasks: one with feedback and one task without feedback. The independent variables are the order in which the tasks are done and whether or not feedback was provided. By using randomised blocks (see Table 1) effects caused by standard order in the experiment are avoided. The learning effect of getting used with the system is tackled by providing a tutorial with a small task before the actual two experiment design tasks. The prototype application provides an environment that is equal for all subjects except for the feedback in word graphs. Consequently, the difference that is measured is the effect of the feedback. In addition, the designs created by each participant are recorded, and the participants fill in a questionnaire at the completion of the session.

1.3 Means of measurement in the experiment

The dependent variables were provided in the prototype's activity log, the word graph acceptance, the ratings of the panel of experts, and the answers of the questionnaire. Below we elaborate on these dependent variables.

1.3.1 Word graph acceptance and activity log

Architects are instructed to accept a word graph during the design process when they find a word graph interesting, or when they think that somehow it contributes to their design. Acceptance of word graphs is made explicit by transferring the word graph from the vertical screen to

Table 1 Tasks and conditions randomly assigned to architects

Conditions, tasks and order	1. Task I 2. Task II	1. Task II 2. Task I
1. With word graphs as feedback 2. Without word graphs as feedback	A	C
1. Without word graphs as feedback 2. With word graphs as feedback	B	C

the Work desk as part of the design draft. The number of accepted word graphs is the measure of applicability of the presented word graphs. The prototype keeps track of the pen activity on the tablet: moving a short distance above the tablet, making strokes on the tablet (sketching and writing), and inserting images or word graphs. We define as no activity a period of 20 s or more that pen activity is not registered. With the information on pen activity an analysis can be made of the duration of the architect's activities.

1.3.2 Assessing by means of a panel of experts

To get insight in the quality of the design results and the design process, a panel of experts from architectural design judges and rates the design results and the design process. By browsing through the process screenshots, each experts can follow the development of the design process. The experts make their ratings individually, without consultation with the other experts. In this way we can rule out consensus judgement by the experts. By having four experts and looking at the mean rating of the panel, personal biases are compensated in the judgements. The experts had no knowledge of the architects who participated in the experiment. The scale used by the panel ranges from 1 to 10, which conforms to general evaluation scores in our setting (1 is the lowest, 10 is the highest score). The panel of experts rates the design quality on four aspects, presented in Table 2. By describing the aspects in more detail, the panel of experts has a common understanding on which aspects should be included in the rating.

Table 2 Design quality aspects

Meeting the requirements of the brief	Function well accomplished of main space. Good relation with the site. Functionality accomplished. User is well considered.
Design process	Many ideas are evaluated. All scale levels are attended in the design process. A design strategy is apparent from the screenshots. There is explicit indication of design problems.
Concept	Completeness. Consistency. Internal logic/reasoning. Level of presentation.
Creativity	Alternative idea(s)/concept(s) made. Breadth of ideas is shown. Interesting, somewhat unusual ideas. Enough complexity in ideas (i.e. depth of the ideas).

1.3.3 Assessing by means of a questionnaire

To obtain the personal experience of the participating architects, we asked them to fill in a questionnaire after the experiment. The questionnaire gives insight into more qualitative measurements of the support offered by the prototype and the effect of words and associations. The architect is given a few definitions beforehand to prevent misunderstandings. Personal data are collected to see whether there is influence on the results depending on age, years of experience, and so forth, and whether an architect had previous experience with a similar design task earlier in his or her career. This information is used to see if there are any effects concerning the degree of homogeneity of the test group.

Architects had to indicate to which degree they agreed to the statement on a scale of seven points [-3 = total disagreement, $+3$ = complete agreement]. Statements concerning creativity are questions on having more associations, ideas, and concepts. Statements concerning fixation are questions on reduction of fixation, enhancement of workflow, pleasure, and on having more unexpected ideas. Although fixation usually is not measured with self-report, we think that architects are just as conscious of the fact if they have a premature commitment to an idea as they are of the fact that they are having more ideas. The questionnaire concludes with open-ended questions and provides room for remarks.

1.4 Experiment participants

About 100 architectural offices in the Netherlands were contacted. In total 18 architects from practice agreed to participate in the experiment. They were promised a small incentive for participation. The architects that are invited to participate in the design experiment all have had at least one year of experience as an architect in practice. The participants were not naïve about the purpose of the experiment. They were requested not to tell anything about the experiment after completion to other architects.

1.5 Experiment procedure

Only one person could use the prototype system at a time. A session took half a day. All 18 sessions could be scheduled within a period of one month. Each participant arrived at the time that his or her appointment had been set for, and was asked to sit down at the ISS prototype design system. He or she got a general introduction to the experiment and a tutorial instruction on how to use the system. The tutorial was the same for each subject. In the tutorial, the system's functionality was explained and the feedback representation in word

graphs clarified. After a short introduction the architect is asked to make a 'dummy' design just for practising with the system and to remove the first learning effects from the experiment. The architects were instructed to insert a word graph when they found it interesting or usable in the design process, or when they had the idea that it influenced the design process. It usually took half an hour to get acquainted with prototype system. After this, they were given the first design task on which they could work one hour, followed by the second design task which also lasts one hour. After the two design sessions the architect is asked to fill in the questionnaire about the use of the system. Each session was video recorded for remote observation during the session, and to provide material for the analysis later.

1.6 Two tasks

The tasks were provided on paper. A plan and additional pictures of the site are included in the paper tasks and they are also available in the design prototype as reference during the design task. The brief includes additional information concerning the background of the task.

1.6.1 Atrium for equality of chances for Eindhoven

One task is called 'Atrium for Equality of Chances for Eindhoven.' The task is to design a building with an atrium in which organisations for equal chances are housed. A site is provided, but the architect may decide for himself where exactly on the site he or she wishes to place the building. The additional information included in the brief explains what organisations should be located within the building. These organisations are described in terms of their goals only.

1.6.2 Bus stop for the Philips high tech campus in Eindhoven

The other task is to design a bus stop for the Philips high tech campus in Eindhoven. The bus stop should radiate Philips' core qualities. The additional information to the brief includes pictures and an explanation on the qualities of the site, and the current state of the campus itself.

2 Results of the experiment

On the basis of the panel of experts, the system's activity log, and the questionnaire, we can study the impact of word graphs in the stimulation of design. The three sources of data are treated in different ways because the nature of these sources is different. The ratings of the panel and the data provided by the system are on a ratio scale. To get an insight into the data from the questionnaire, however, it is assumed that the scores are measured on an interval scale. Two methods are used to assess the data statistically: comparing means with a two-tailed *t*-test

and comparing means with a paired-samples *t*-test. The outcomes of the different sources, the design draft from the screenshots, and the remarks made in the questionnaire are considered together to formulate the conclusions.

Gender (male and female), years of experience (ranging from 1 to 36 years), and attitude towards writing in the early phase of the design process (using words in practice and liking words) were variables that are considered to have an effect on the results. Results are presented per means of measurement.

2.1 Word graph acceptance and activity log

The measure for the graphs' usefulness as design content is inferred from the number of accepted graphs (Table 3). Based on our instruction to the architects to accept a word graph when it is found interesting and when it contributes to the design process, we assumed that word graphs would be accepted when something new is found in the word graphs. In other words, acceptance of word graphs would indicate novelty and provide proof of enhanced creativity. The graph acceptance ranges from 1 to 19 word graphs per session (max. 1 h) with feedback, which is promising. As can be seen in Table 3, the behaviour in accepting word graphs was rather diverse among architects. We can also observe that a fragment of the total amount of generated word graphs was accepted by architects.

In the questionnaire, some architects indicated that acceptance of word graphs does not necessarily imply novelty. We also found that both

Table 3 Word graph acceptance by architects ($n = 18$)

Architect	1	2	3	4	5	6	7	8	9
Task done with feedback in word graphs	B	B	B	B	A	A	B	B	B
Number of written words in task wfb	60	51	76	34	57	72	47	51	60
Number of accepted word graphs (wfb)	15	7	10	5	1	2	2	12	6
Total number of word graphs generated during the task (wfb)	210	212	265	91	123	252	158	156	158
Acceptance percentage of word graphs (%) (wfb)	7.1	3.3	3.8	5.5	0.8	0.8	1.3	7.7	3.8
Architect	10	11	12	13	14	15	16	17	18
Task done with feedback in word graphs	B	A	B	A	A	A	B	A	A
Number of written words in task wfb	33	26	20	23	31	36	51	37	106
Number of accepted word graphs (wfb)	6	10	3	3	2	1	19	8	18
Total number of word graphs generated during the task (wfb)	151	101	45	73	92	74	103	90	315
Acceptance percentage of word graphs (%) (wfb)	4.0	9.9	6.7	4.1	2.2	1.4	18.4	8.9	5.7

wfb, with feedback in word graphs; A, Atrium; B, Bus stop.

architects who indicated in the questionnaire a positive attitude or a negative attitude towards the system, accepted word graphs. However, it can be stated that architects accepted more word graphs when they indicated in the questionnaire that they were used to working with words while designing.

From comparing the creation log of generated word graphs with the written words in the design draft (Figure 5), we can see that a short while after word graphs are shown on the vertical screen, architects also write down a number of these generated words that occurred earlier in the word graphs. We may infer that although instructed to explicitly accept word graphs and place them in the design draft, the architects had forgotten to do so. This implies that the architect has used more of the feedback in word graphs than the graph acceptance by itself indicates.

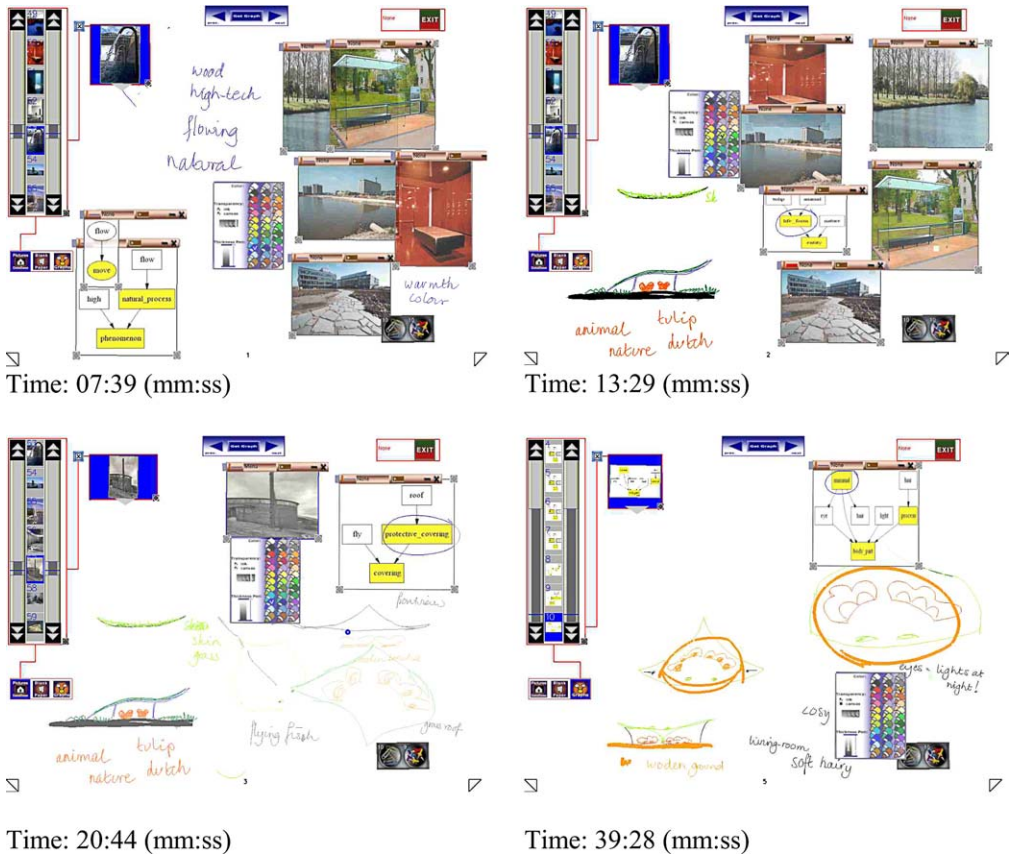

Figure 5 Four examples (out of 24) of design draft screenshots during the design process of the bus stop task

The question whether activity increases or decreases when the architect is given feedback in word graphs, has bearing on how useful the feedback is for keeping the design process going. Figure 6 shows the periods of pen activity and acceptance of words graphs in the design draft of one participating architect. The percentage of the duration of no activity and activity gives insight of the amount of no activity relatively per session (Table 4). The total duration of the task (minutes), the total amount of no activity, and the percentage of the duration of no activity and activity play a role in the assessment of design stimulation. This is so because if the task becomes more difficult, the design takes longer to finish and more time is spent of doing nothing. Mean_1 is the mean duration of all architects doing a task with feedback in word graphs and Mean_2 is the mean duration of all architects doing a task without feedback in word graphs. The standard deviation is given with both means. The means are compared to be significantly different. It is found significant, on a 5% probability level, that giving feedback is of influence on the activity in the design process. With feedback the architects have longer periods of both activity and of no activity. The percentage of the duration of no activity and the duration of the session also increased when given feedback. It can be concluded that the feedback is not reducing the duration of the task or the duration of no activity, according to the measures of the prototype's activity log. In the sense of the duration of activity and no activity, the system does not stimulate the design.

2.2 *Panel of experts*

The panel of experts rated the design (process) on the basis of the design draft as displayed in Figure 5. In Table 5, the mean ratings of the panel of experts are displayed. Mean_1 is the mean of data related to ratings of the tasks done with feedback in word graphs, while Mean_2 is the mean of data related to ratings of the tasks done without feedback in word graphs. The means are compared to be significantly different. A general observation indicates that the tasks with feedback are rated with a lower grade (differences vary between 0.25 and 0.5), but these differences between the mean ratings of the panel of experts are not proved to be significant. This leads to the conclusion that, according to the panel of experts, the architect was not helped in meeting the requirements of the

Figure 6 Activity log (filled blocks indicate activity, white indicates no activity) and accepted word graphs (lines), architect 8

Table 4 Compare means of the duration of the design activity and no activity – paired-samples *t*-test

	$Mean_1$ ($n = 18$) wfb	Std. $Dev._1$	$Mean_2$ ($n = 18$) nfb	Std. $dev._2$	Mean diff.	Sig. (two-tailed)
Duration no activity	7.50	2.48	5.17	3.84	−2.33**	0.016**
Duration activity	51.61	4.88	47.94	5.82	−3.67**	0.050**
Percentage no activity/duration total	12.6	4.0	9.1	5.4	−3.5**	0.010**

**Significant results 'Sig. $\leq$ 0.05' wfb: with feedback in word graphs. *Significant results 'Sig. $\leq$ 0.10' nfb: no feedback in word graphs.

task, having a better design process, building a concept, or in being more creative by the system's feedback, nor the opposite.

In closer observation, it seems that the attitude to using words in practice influences the outcome (see Table 6). $Mean_1$ is the mean rating of 12 architects who have indicated to use words in the early phase of the design process in practice and $Mean_2$ is the mean rating of six architects who have indicated not to use words in the early phase of the design process in practice. Architects who indicated in the questionnaire that they are used to using words in the early phase of the design process in their daily practice, showed a higher level of creativity according to the panel of experts when the task was done without word graph feedback. The difference is 0.47 and significant on a 5% probability level when equal variances are not assumed. However, when the task was done with word graph feedback this effect was not measured. We have no data to explain this phenomenon.

Previous experience in making the task did have some influence on the quality of the design as rated by the panel of experts (see Table 7). $Mean_1$ is the mean rating of 12 architects who have indicated not to have done a task like the ones in the experiment, while $Mean_2$ is the mean rating of six architects who have indicated to have done a task like (the) one(s) in the experiment. Having done one or both tasks in a similar task before influences the design process ratings negatively when no feedback is given. It may be the case that architects remembered what they have

Table 5 Compare means of ratings given by the panel of experts – paired-samples *t*-test

	$Mean_1$ ($n = 18$) wfb	Std. $dev._1$	$Mean_2$ ($n = 18$) nfb	Std. $dev._2$	Mean difference	Significance (two-tailed)
Quality	6.25	0.66	6.59	0.51	−0.34	0.165
Design process	6.21	0.46	6.45	0.47	−0.24	0.190
Concept	6.29	0.83	6.74	0.54	−0.45	0.130
Creativity	6.37	0.92	6.67	0.55	−0.30	0.353

**Significant results 'Sig. $\leq$ 0.05'.*Significant results 'Sig. $\leq$ 0.10'. wfb: with feedback in word graphs. nfb: no feedback in word graphs.

Table 6 Compare means – independent samples *t*-test, statistical results for influence 'word usage' – ratings for tasks with feedback/without feedback

	$Mean_1$ ($N = 12$)	Std. $dev._1$	$Mean_2$ ($N = 6$)	Std. $dev._2$	Mean diff.	Sig. A	Sig. B
Quality (wfb)	6.16	0.75	6.44	0.47	−0.28	0.414	0.343
Design process (wfb)	6.12	0.50	6.40	0.33	−0.28	0.230	0.174
Concept (wfb)	6.18	0.98	6.50	0.40	−0.32	0.458	0.342
Creativity (wfb)	6.29	1.13	6.52	0.22	−0.23	0.633	0.509
Quality (nfb)	6.71	0.45	6.35	0.59	0.35	0.172	0.230
Design process (nfb)	6.49	0.50	6.38	0.45	0.12	0.643	0.632
Concept (nfb)	6.87	0.56	6.48	0.41	0.39	0.155	0.119
Creativity (nfb)	6.82	0.59	6.35	0.30	0.47**	0.089*	0.040**

Sig. A = significance two-tailed equal variances assumed, Sig. B = significance two-tailed equal variances not assumed. **Significant results 'Sig. $\leq$ 0.05', *significant results 'Sig. $\leq$ 0.10'. wfb: with feedback in word graphs, nfb: no feedback in word graphs.

made and skipped some parts in the design process, with the consequence that their design draft became less clear to the panel of experts. Possibly, architects were less motivated when a similar task has already been done in the past.

2.3 Questionnaire

In assessing the role of words and associations in stimulating design, the opinions of the architects must be made explicit as well. The answers to the questions concerning the (dis)agreement on statements were analysed by a comparing means method with a two-tailed *t*-test. The null hypothesis is that the architects answered neutral ($\mu_0 = 0$) and the alternative hypothesis expresses either a disagreement or an agreement, where μ_1 is 1 or −1. When a result is found significant a second null hypothesis is defined which checks whether $\mu_0 = 1$ and the alternative

Table 7 Compare means – independent samples *t*-test, statistical results for influence 'experience in task' – ratings for tasks with feedback/without feedback

	$Mean_1$ ($N = 12$)	Std. $dev._1$	$Mean_2$ ($N = 6$)	Std. $dev._2$	Mean diff.	Sig. A	Sig. B
Quality (wfb)	6.29	0.74	6.17	0.53	0.13	0.719	0.688
Design process (wfb)	6.25	0.52	6.13	0.33	0.13	0.600	0.542
Concept (wfb)	6.39	0.80	6.08	0.92	0.31	0.478	0.505
Creativity (wfb)	6.48	0.89	6.15	1.02	0.33	0.486	0.513
Quality (nfb)	6.72	0.48	6.33	0.52	0.39	0.135	0.159
Design process (nfb)	6.64	0.48	6.08	0.13	0.55**	0.015**	0.002**
Concept (nfb)	6.71	0.53	6.79	0.60	−0.08	0.766	0.778
Creativity (nfb)	6.76	0.59	6.48	0.46	0.28	0.322	0.288

Sig. A = significance two-tailed equal variances assumed, Sig. B = significance two-tailed equal variances not assumed. **Significant results 'Sig. $\leq$ 0.05', *significant results 'Sig. $\leq$ 0.10'. wfb: with feedback in word graphs, nfb: no feedback in word graphs.

hypothesis would mean that μ_1 is larger than 1. Here a one-tailed t-test was used.

In assessing whether the architects are neutral towards the importance of words or not, we performed a two-tailed t-test. The mean is 1.61, which is significant according to the testnumber of 4.334 (see Table 8). We can conclude that architects tend to agree that words are important for them in the early phase of the design process to express their ideas. When comparing with Table 3, we can see that not all architects use written words in the same amount, but comparing the answers from the questionnaire with the design draft, architects appear to use written words more than they are aware of.

The influence of feedback in word graphs on the design in terms of having more associations, (unexpected) ideas, and concepts as indicated in the questionnaire are presented in Table 9. All means are above zero. The increase of associations has a t-statistic of 5.497, which is significant beyond the 5% probability level. The t-statistic where $\mu_0 = 1$ is 1.374 is significant for a probability level of 10%. The increase of number of ideas has a t-statistic of 2.650, which is significant beyond the 5% probability level. The increase of number of concepts has a t-statistic of 3.500, which is significant beyond the 5% probability level. The architects tend to agree to have more unexpected ideas, when stimulated by the system (t-statistic of 3.063, significant beyond the 5% probability level). We can conclude that architects tend to agree with the statement in the questionnaire that the presentation of word graphs provides more associations, (unexpected) ideas, and concepts. Therefore, architects experience mostly that they have more associations with the system.

The effect of word graphs shown in Table 10 concerns pleasure, fixation, and process. The mean for fixation is 0.50, which is not significant according to the testnumber of 1.534. Some architects experienced a reduction in fixation, some did not, and most were neutral. The mean for pleasure is 0.94 and for process it is 1.00, with t-statistics respectively 2.521 and 2.699, which both are significant beyond the 5% probability

Table 8 Importance of words in expressing ideas in the early phase of the design process − statistics ($n = 18$, scale [−3,3])

Mean	Standard deviation	t-value, $\mu_0 = 0$, two-tailed	t-value, $\mu_0 = 1$, one-tailed
1.61	1.58	4.334**	1.644*

One-tailed, critical testnumbers: 1.333 (10% probability level*) and 1.740 (5% probability level**). Two-tailed, critical testnumbers: 1.740 (10% probability level*) and 2.110 (5% probability level**).

Table 9 Agreement on statement: Working with the ISS, I came up with more 'X' myself while designing, than without the stimulation of the system − statistics ($N = 18$, scale [−3,3])

'X'	Mean	Standard deviation	t-value, $\mu_0 = 0$, two-tailed	t-value, $\mu_0 = 1$, one-tailed
Associations	1.33	1.03	5.497**	1.374*
Ideas	0.61	0.98	2.650**	−1.686
Concepts	0.78	0.94	3.500**	−1.000
Unexpected ideas	0.89	1.23	3.063**	−0.383

One-tailed, critical testnumbers: 1.333 (10% probability level*) and 1.740 (5% probability level**). Two-tailed, critical testnumbers: 1.740 (10% probability level*) and 2.110 (5% probability level**).

level. We can conclude that architects tend to agree that there is an improvement of pleasure and the process and workflow when using the system. Based on the questionnaire, there is no evidence of a reduction of fixation.

We analysed whether or not there were learning and order effects. Although we used randomised blocks in the design of the experiment, we still measured that there are learning and order effects in the tasks.

In Table 11 the influence of which task is done with feedback is displayed, where Mean_1 is the mean of the group of architects who have done the atrium task with feedback, and Mean_2 of the architects who have done the bus stop task with feedback. The bus stop task with word graph feedback showed reduced fixation (1.10 on a 7 point scale, 5% probability level), while the atrium task showed increased fixation (−0.25 on a 7 point scale, 5% probability level) with word graph feedback. This finding might indicate that the atrium task was too complex by itself, or that the bus stop task is less abstract and therefore leads faster to useful associations that the system can show.

In Table 12, an overview is given on the influence of doing the task with feedback in word graphs first. Here, Mean_1 stands for the mean of the

Table 10 Agreement on statement: the use of the Idea Space System X − statistics ($N = 18$), scale [−3,3]

'X'	Mean	Standard deviation	t-value, $\mu_0 = 0$, two-tailed	t-value, $\mu_0 = 1$, one-tailed
Gave more pleasure	0.94	1.59	2.521**	−0.148
Reduced fixation	0.50	1.38	1.534	
Gave a better process and workflow	1.00	1.57	2.699**	0.000

One-tailed, critical testnumbers: 1.333 (10% probability level*) and 1.740 (5% probability level**). Two-tailed, critical testnumbers: 1.740 (10% probability level*) and 2.110 (5% probability level**).

Table 11 Compare means – independent samples *t*-test, statistical results for influence 'task done with feedback' – statistics ($N = 18$), scale [−3,3]

	$Mean_1$ ($N = 8$) wfb1	Std. $dev._1$	$Mean_2$ ($N = 10$) wfb2	Std. $dev._4$	Mean diff.	Sig. A	Sig. B
Fixation reduced	−0.25	1.49	1.10	0.99	−1.35**	0.035**	0.048**
Better workflow	1.13	1.36	0.90	1.79	0.23	0.773	0.766
More pleasure	1.00	1.51	0.90	1.73	0.10	0.899	0.898

Sig. A = significance two-tailed equal variances assumed, Sig. B = significance two-tailed equal variances not assumed. **Significant results 'Sig. ≤ 0.05', *significant results 'Sig. ≤ 0.10'. wfb1: first task was with feedback in word graphs, wfb2: first task was with feedback in word graphs.

group of architects who have done the task with feedback first, while $Mean_2$ has done the task with feedback as the second task. More pleasure was experienced with the use of the system when the task with feedback was done second (with a difference of 1.48 on a 7 point scale, when the two-tailed variances are assumed to be equal, the significance of 0.047 is beyond the 5% probability level). This leads us to expect that increased experience with the system may lead to improved pleasure working with the system.

Table 13 shows the influence of the architects' attitude towards words. $Mean_1$ stands for the mean of the group of architects who use words in the early phase of the design process in practice and $Mean_2$ stands for the mean of the group of architects who do not. It has been proved significant that the architects who normally use words in the early phase of the design process experience more pleasure and a better workflow in using the system than those architects whom attribute no specific value to words (the differences are 1.67 and 2.00, with a significance beyond the 5% probability level).

There were no effects in terms of more associations, (unexpected) ideas, and concepts.

Table 12 Compare means – independent samples *t*-test, statistical results for influence 'which task done first' – statistics ($N = 18$), scale [−3,3]

	$Mean_1$ ($N = 8$)	Std. $dev._1$	$Mean_2$ ($N = 10$)	Std. $dev._2$	Mean diff.	Sig. A	Sig. B
Fixation reduced	0.75	1.49	0.30	1.34	0.45	0.509	0.516
Better workflow	0.38	1.92	1.50	1.08	−1.13	0.135	0.169
More pleasure	0.13	1.96	1.60	0.84	−1.48**	0.047**	0.078*

Sig. A = significance two-tailed equal variances assumed, Sig. B = significance two-tailed equal variances not assumed. **Significant results 'Sig. ≤ 0.05', *significant results 'Sig. ≤ 0.10'. wfb: with feedback in word graphs, nfb: no feedback in word graphs.

Table 13 Compare means – independent samples *t*-test, statistical results for influence 'attitude towards words' – statistics ($N = 18$), scale [−3,3]

	$Mean_1$ ($N = 12$)	Std. $dev._1$	$Mean_2$ ($N = 6$)	Std. $dev._2$	Mean diff.	Sig. A	Sig. B
Fixation reduced	0.67	1.61	0.17	0.75	0.50	0.486	0.384
Better workflow	1.67	1.23	−0.33	1.37	2.00**	0.006**	0.014**
More pleasure	1.50	1.00	−0.17	2.04	1.67**	0.031**	0.106

Sig. A = significance two-tailed equal variances assumed, Sig. B = significance two-tailed equal variances not assumed. **Significant results 'Sig. ≤ 0.05', *significant results 'Sig. ≤ 0.10'. wfb: with feedback in word graphs, nfb: no feedback in word graphs.

There are some remarks of the architects that are valuable concerning further pursuing continuation on words and associations and word graphs in CAAD. Architects appreciated being helped with explaining or writing about their design or ideas, to receive hints at a direction of thinking, abstraction, overview, and to see logical structures in words that express design ideas. One architect stated: 'when I don't know what to do anymore for a while, I just write down some words and who knows what comes of it? It is an extra possibility for me next to watching the images to get inspiration. I had a more constant flow of ideas. It is pleasant to work with the feeling more is coming out of me.' Most opinions about ISS are positive, but there are five architects who think there were too many word graphs generated as feedback. They had a problem with the fact that there was not an option to reduce the number of word graphs, select the words that should not or should be in the word graphs. On top of that, architects often found the words too abstract, and three architects considered words not important to express their ideas in the design process and had therefore less benefit from the system.

3 *Conclusion*

Based on the analysis above, we can draw conclusions about the use of word graphs and their impact on design, as well as about the research method used in this project.

3.1 *The use of words in architectural design*

We found that there is a remarkable difference between the actual amount of annotations (word count of the recorded design drafts) and the architects' own assessment of this amount (as stated in the questionnaire). Architects write more than they think they do. Despite this fact, the determining factor whether support via words is useful seems to depend on the architect's predisposition to words. The ISS therefore, is a good aid to make architects aware of their own writing.

Hypernym- and hyponym-relationships are often too abstract and therefore more likely to be useless. The large amount of word graphs is helpful in a divergent, explorative thinking process, when it is important that an idea leads to many different ideas, rather than as a means to quickly reach focus and convergence. This implies that the system has more use in explorative phases of design than in convergent phases of design. Further research is needed to determine which semantic associations are interesting and which are not, which (semantic) associations may stimulate convergent thinking, and how filters can be defined to filter out non-interesting feedback for the architect.

Despite the fact that architects were using the prototype design system for the first time, it appeared that this did not significantly influence the design negatively, except for the duration of no activity and of making the task. For most architects, the system did not aid in dealing with complexity, with the exception of two architects who indicated in the questionnaire that the system helped in dealing with complexity. There is no proof of significant reduction of fixation or improvement of creativity.

Stimulation was found mostly on the pleasure in using the word graphs, an improvement of the workflow, in having more (unexpected) ideas and concepts, and especially in having more associations. The aspect of unexpectedness implies that there is useful novelty generated by the system, which may lead to increased creativity in the design. Therefore, we feel justified to predict a positive effect when the prototype will be developed further.

We observed that architects wrote down words that were generated by the system earlier, even when they did not explicitly accept those word graphs. This underlines that the architects find the information in the word graphs interesting or useful. Furthermore, we observed that architects who indicated to find words important and to use them in design practice also indicated to be more distracted by the feedback in word graphs. We expect there are two kinds of architects: language-sensitive, and language non-sensitive architects. The first group will find more meaning in words and tend to play more with words and their meanings. If this is true, then prolonged experience with the system will not remove the learning effect.

3.2 Research method

The use of a working prototype enables us to study architects as they make a design, while the feedback in word graphs appears in real-time, and with automatic recording of significant aspects through an activity

log. In this research we have developed and used three sources to obtain data: the reactions to the system expressed in a questionnaire, a panel of experts to rate the designs, and the recording and measurements via the prototype activity log. From the discussion above, we can conclude that none of these three sources are by themselves adequate and sufficient, but that they need to be considered in concert to understand what the impact of word graphs to the architect is.

The outcome as a whole is not as positive as we hoped for, but there may be some aspects related to the research method, the experiment, and the prototype that may have played a role.

There are some aspects related to the participants that may have influenced the outcome. Firstly, an experiment like this may have produced more subtle answers to the research questions when more than 18 architects would have participated in the experiment. Secondly, one of the architects did not finish his task. Unfortunately this architect was part of block A (see Table 1), so it turned out that only three architects instead of four completed that particular series of tasks. All other blocks (B, C, and D) contained five architects. This may have had some impact on the data from the experiment. Thirdly, it is possible that inter-subjective analysis could give interesting results, but we did not make this analysis.

Although the answer to questions like 'what is a better idea or design?' or 'what is creativity?' can be researched by means of a panel of experts, the explanation of their judgement or a definition of a precise measure for quality is yet another issue. By having multiple experts rate the design two problems are evaded: to a certain extent the subjective opinion of the researcher, and the need to define in a measurable way what a good design is or what creativity is. Multiple experts in a panel may balance out different personal views, and by using experienced 'raters' consistency in the judgement may be expected. Therefore we have to consider the mean rating of the panel as the measure that is considered valid, and not the separate ratings. On the other hand, it can also be argued that by its very nature, the system is a personal one and those capable of judging whether the system has supported creativity are the users of the system themselves.

The architects' opinion on the system assessed with the questionnaire is rather positive. It must be kept in mind, however, that the danger with a questionnaire is that the architects may have given social-desirable responses: they may have a tendency to give an answer that is in accordance with what is the desirable outcome of the experiment. Further

there may be some bias concerning the set-up of the questionnaire, due to for instance the use of a seven-scale answering possibility: the architects could not give an answer that lies in between two answers.

As stated in Section 2, not just the experiment, but also the implementation of the prototype triggers questions about the viability of the proposed research method. From a pragmatic point of view, choices need to be made in designing and implementing the prototype. There are three choices that have influence on the experiment: (i) implementation of only semantic associations and not including other types of associations, (ii) the display of word associations as graphs, and (iii) implementation of ISS with a natural and intuitive interface and the feedback on a vertical screen. The choice to present the word graphs on the vertical screen, separate from the working area, may have caused some problems. Possibly, when the architect is very much focused on the design, he or she may forget to take a look at the feedback. It will take more development and testing to see how 'close' or 'direct' the feedback can be presented to the architect without creating a sense of crowding the work area or disrupting the flow of thought. From the experiments it appeared that the ISS platform still took more learning time than was expected from earlier experiments. This should be easily solved by having the architect practice longer with the system, but it also illustrates our point that expectations about potential utilities of various design support tools can only be tested in the more complex, everyday design activity, despite the additional methodological questions it raises.

From the experiment and the results we may conclude that already the provision of a natural language ontology to generate associations can be a good aid for architects to generate more ideas during the design process. We have also seen, however, that this mainly applies for those architects who have a positive disposition to the use of words. Other types of association, such as visual associations or personal associations (also known as episodic) have not been investigated. It may well be that these types of association are fruitful for other 'types' of architects. Based on our current findings, however, we can caution against overly generalising expectations: it may well be that such functionality works well only for those architects who positively view such aspects.

Taking in account the remarks of the architects, extra filtering by the system or by the architects themselves must be possible to limit the amount of graphs and the level of abstraction in the ISS. In this way better results can be expected.

4 Future work

As stated above, semantic associations are not the only types of associations. It is very likely that other types of associations, such as visual associations, episodic associations, and so forth play an important role in constructing the design rationale. The acquisition of such structures is quite difficult because of the personal nature. It will require long periods of tracking an individual designer to establish just the beginning of such elaborate personal association structures. In terms of the design draft, at least one element is available to infer such relations: mark making in the form of arrows, lines, encircling, and so forth. Presumably if marks in relation to the rest of the design draft are better understood, or if the right set of marks is agreed upon to express certain episodic associations without loosing intuitiveness, marks can extend the understanding of the design draft in a more personal way.

Acknowledgements

The authors would like to thank all architects who participated in the experiment. Thanks are also due for Dima Aliakseyue (Industrial Design Department of the Eindhoven University of Technology) and Joran Jessurun (Design Systems Group of the Eindhoven University of Technology) who together made it possible to run the ISS in the Design Laboratory. We wish to thank Theo Arentze for his rigorous comments and advice.

References

Achten, H H (1997) Generic representations – an approach for modelling procedural and declarative knowledge of building types in architectural design, Ph.D. thesis, Technische Universiteit Eindhoven, the Netherlands

Alexander, C (1964) *Notes on the synthesis of form* Oxford University Press

Aliakseyeu, D, Martens, J B and Subramanian, S, et al (2001) Visual interaction platform, in *Proceedings of Interact 2001*, Tokyo pp 680–681

Amabile, T M (1983) *The social psychology of creativity* Springer Verlag, New York

Amabile, T M (1996) *Creativity in context* Westview Press, Boulder, CO

Archer, L B (1969) The structure of the design process, in **G Broadbent and A Ward** (eds) *Design methods of architecture*, Witteborn, New York

Christiaans, H (1992) *Creativity in design: the role of domain knowledge in designing* Lemma, Utrecht

Cordingly, E S (1989) Knowledge elicitation techniques for knowledge-based systems, in **D Diaper** (ed.) *Knowledge elicitation: principles, techniques and applications*, Ellis Horwood

Do, E Y-L (2002) Drawing marks, acts, and reacts: toward a computational sketching interface for architectural design *Artificial Intelligence for Engineering Design Analysis and Manufacturing* Vol 16 No 3 pp 149–171

Dorst, C H (1997) Describing design − a comparison of paradigms, Ph.D. thesis, Technical Universiteit Delft, the Netherlands
Ericsson, K A and Simon, H A (1993) *Protocol analysis* MIT Press, Cambridge, MA
Finke, R A, Ward, T B and Smith, S M (1992) *Creative cognition − theory, research, and applications* MIT Press, Cambridge, Massachusetts
Goldschmidt, G (1991) The dialectics of sketching *Creativity Research Journal* Vol 4 pp 123−143
Koutsofios and North, S C (1993) Drawing graphs with dot, Technical report, AT&T Bell Laboratories, Murray Hill, NJ
Lawson, B and Loke, S M (1997) Computers, words and pictures *Design Studies* Vol 18 No 2 pp 171−183
van der Lugt, R (2001) Sketching in design idea generation meetings, Ph.D. thesis, Technische Universiteit Delft, the Netherlands.
Miller, G A, Beckwith, R and Fellbaum, C (1990) Introduction to WordNet: an on-line Lexical database *International Journal of Lexicography* Vol 3 No 4 pp 235−244
Purcell, T A and Gero, J S (1996) Design and other types of fixation *Design Studies* Vol 7 No 4 pp 363−383
Reymen, I M M J (2001) Improving design processes through structured reflection: a domain-independent approach. Ph.D. thesis, Technische Universiteit Eindhoven, the Netherlands
Rowe, P G (1987) *Design thinking* MIT Press, Cambridge, Massachusetts
Schön, D (1983) *The reflective practitioner* Basic Books, New York
Segers, N M (2004) Computational representations of words and associations in architectural design − development of a system supporting creative design, Ph.D. thesis, Technische Universiteit Eindhoven, the Netherlands
Segers, N M and de Vries, B (2003) The idea space system − words as handles to a comprehensive data structure, in **Mao-Lin Chiu**, **Jin-Yeu Tsou**, **Thomas Kvan**, **Mitsuo Morozumi and Tay-Sheng Jeng** (eds) *Digital Design − Research and Practice− Proceedings of the 10th International Conference on Computer Aided Architectural Design Futures*, Kluwer Academic Publishers, Dordrecht pp 31−40
Silberman, Y, Miikkulainen, R and Bentin, S (2001) Semantic effect on episodic associations, in **J D Moore and K Stenning** (eds) *Proceedings 23rd Annual Conference of the Cognitive Science Society*, Cognitive Science Society, Edinburgh
Simon, H (1973) The ill structure of ill-structured problems *Artificial Intelligence* Vol 4 pp 181−204
Suwa, M, Gero, J S and Purcell, T (1998) The roles of sketches in early conceptual design processes, in *Proceedings of Twentieth Annual Meeting of the Cognitive Science Society*, Lawrence Erlbaum, Hillsdale, New Jersey pp 1043−1048
Valkenburg, R C (2000) The reflective practice in product design teams, Ph.D. thesis, Technische Universiteit Delft, the Netherlands